Cognition Helps Vision: Recognizing Biological Motion Using Invariant Dynamic Cues

Nicoletta Noceti[1(✉)], Alessandra Sciutti[2], and Giulio Sandini[2]

[1] DIBRIS, Università di Genova, Genova, Italy
nicoletta.noceti@unige.it
[2] RBCS, Istituto Italiano di Tecnologia, Genova, Italy
{alessandra.sciutti,giulio.sandini}@iit.it

Abstract. This paper considers the problem of designing computational models of the primitives that are at the basis of the visual perception of motion in humans. The main contribution of this work is to establish a connection between cognitive science observations and empirical computational modeling. We take inspiration from the very first stage of the human development, and address the problem of understanding the presence of biological motion in the scene. To this end, we investigate the use of coarse motion descriptors composed by low-level features inspired by the Two-Thirds Power Law. In the experimental analysis, we first discuss the validity of the Two-Thirds Power Law in the context of video analysis, where, to the best of our knowledge, it has not found application so far. Second, we show a preliminary investigation on the use of a very simple motion model for characterizing biological motion with respect to non-biological dynamic events.

1 Introduction

The interactions with other people or with the surrounding environment are easy and natural tasks for human beings, triggered by an innate predisposition. Nevertheless, it is well accepted in the cognitive science community that a mature social awareness is subject to the acquisition of a sequence of temporally-ordered perceptual and social skills, going from the detection of target of potential interest [10], to the capability of inferring the intentions of other people and the goals of their actions [8].

This work considers the development of visual perception capabilities in humans, and tries to establish a connection between the observations coming from the cognitive science world and the computational modeling side. The long-term goal of our research is the design of computational vision models able to replicate on an artificial system the developmental stages of motion perception in humans. This is of particular interest, for instance, in the robotics field, where the design of methods for a natural human-robot interaction is one of the great challenges of the research nowadays.

In this paper we specifically refer to the earliest stages of human development, and consider in particular the capability of understanding the presence

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 676–686, 2015.
DOI: 10.1007/978-3-319-23234-8_62

of biological motion in the surrounding environment, a skill humans, and not only, exhibit early after birth [20]. This ability triggers the development of social interaction, since it allows the detection of potential interaction partners in the scene.

We consider a binary classification setting in which characterizing biological movements with respect to non-biological dynamic events. As for the first class, we are particularly interested in sequences of human actions typical of interactions, as repositioning objects or pointing towards a certain 3D location.

Given a video stream, we initially detect the regions where the motion is occurring using the optical flow, then we extract a set of low-level features inspired by the *Two-Thirds Power Law*, which has been experimentally proved to be an invariant property of biological motion, and human movements in particular [18,23,24,26]. We adopt a coarse motion representation leveraging on the fact that if humans show a predisposition for biological motion right after birth, when the amount of visual information is still very limited, then *it is likely* that it may depend on very simple motion information.

We consider two different levels of compression of such information over time – computing a point-based and a region-based descriptor – and evaluate their use with binary SVMs classifiers equipped with appropriate Multi-Cue kernels [21].

Related Works. Since we are primarily interested in capturing abilities typical of the early months of human development, we do not address classical action recognition tasks (very fertile disciplines in fields as video surveillance, video retrieval and robotics [4,16,28]), abilities which are likely to be gained at later stages of development, also thanks to the infants' prior motor experience [3]. Within this contexts, an approach sharing similarities with our work is [19] where the authors consider the problem of biological motion classification using joints trajectories. However, they refer to the characterization of a single class of human motion (walking) with respect to others (as boxing or jumping).

Instead, works on human perception of biological motion can be traditionally found in the field of cognitive science, where particular interest has been posed on the relative importance of visual features that are (presumably) at the basis of this strong ability [1,7,22]. In most of such works *point-light displays* or motion caption systems are adopted.

The Two-Thirds Power Law has been related to the motion perception of humans [6,24,26], and it is considered a well-known invariant property of human movements [12,18,25]. Its applicability has been empirically verified mostly for upper-limb movements, but also for eye motion [27], locomotion [23], and to the purpose of movement prediction [9]. The relation between motion and the quantities involved in the law has been also deeply analysed [12,25]. In [13] the authors show that white Gaussian noise also obeys this power-law.

To the best of our knowledge, this is the first attempt of applying the Two-Thirds Power Law in the context of video analysis, on data measured from video stream and thus, by construction, less controlled. Also, with respect to previous works, we consider a broader range of possible human movements.

The remainder of the paper is organized as follows. In Sec. 2 we briefly review the theory of the Two-Thirds Power Law, which is used as an inspiration to introduce the low-level features we consider, in Sec. 3. Sec. 4 describes the motion representation we adopt and sets the scene for the learning problem. We report the experimental analysis in Sec. 5 and we leave the final discussion to Sec. 6.

2 The Two-Thirds Power Law

Each dynamic physical event can be easily described by its spatial trajectory – which defines the *shape* of the motion – as well as many other quantities – as the evolution of length, velocity or direction. All of them represent evidences of the dynamics, and are in general interconnected with each other.

For the specific case of human motion, it is acknowledged the validity of an exponential relation between functions measured from the motion [6,23,26]. The relation can be formulated as

$$V(t) = K(t) \left(\frac{R(t)}{1 + \alpha(t) R(t)} \right)^{\beta} \tag{1}$$

where $V(t)$ is the tangential velocity, $R(t)$ is the radius of curvature, $\alpha(t) \geq 0$ depends on the average motion velocity (and is null in absence of points of inflection in the trajectory), $K(t) \geq 0$, depends on tempo and length of the motion [25]. In case $\alpha(t) = 0$ the law can be written in the alternative, yet equivalent, form

$$A(t) = K(t) C(t)^{1-\beta} \tag{2}$$

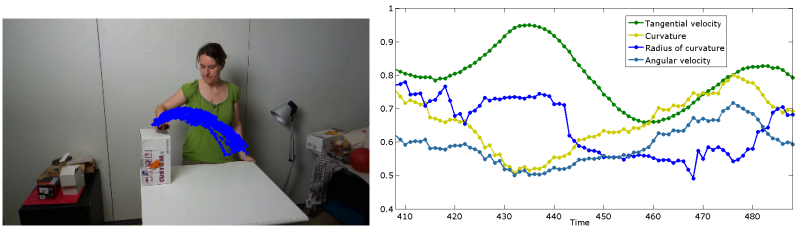
where $A(t) = \frac{V(t)}{R(t)}$ and $C(t) = \frac{1}{R(t)}$. In adults, the value of β (estimated most often for drawing movements) is very close to $\frac{1}{3}$, and so the law in Eq. 2 is usually referred to as *Two-Thirds Power Law*.

Although this relation has been deeply investigated in the fields of human motion perception analysis and cognitive science, the application in the context of artificial intelligence and computer vision is still unexplored. In the following, thus, we consider the use of a motion descriptor guided by the law and discuss its adoption in a video analysis setting.

3 From the Law to the Features

Inspired by the Two-Thirds Power Law, our idea is to describe an observed motion with a vector of low-level spatio-temporal features, computational counterparts of the variables involved in the mathematical formulation.

At each time instant t, we start by evaluating the optical flow with a dense approach (as [5]) and detecting the regions of interest $\mathcal{R}(t)$ – i.e. the regions where the motion is occurring – with a hysteresis thresholding on the magnitude. Notice that, in general, at each time instant we may detect more than one



(a) (b)

Fig. 1. Left: an example of the trajectory of a point describing the dynamic of a sequence of lifting actions. Right: the temporal series of the low-level features we computed on a sub-part of the sequence.

region. They can correspond to different portions of a single common event (e.g. when gesticulating with both hands), or they may indicate the co-occurrence of multiple events.

We then associate each point $\mathbf{p}_i(t) \in \mathcal{R}(t)$ with a feature vector

$$\mathcal{F}(\mathbf{p}_i(t)) = [\hat{V}_i(t), \hat{C}_i(t), \hat{R}_i(t), \hat{A}_i(t)] \tag{3}$$

where the features denote, respectively, tangential velocity, curvature, radius of curvature and angular velocity estimated for the point as follows. Let $(u_i(t), v_i(t))$ be the optical flow components. We define the spatio-temporal velocity of the point as $\hat{\mathbf{V}}(t) = (u_i(t), v_i(t), \Delta_t)$, where Δ_t is the temporal displacement between observations of two adjacent time instants. The velocity magnitude is computed as $\hat{V}(t) = \sqrt{u_i(t)^2 + v_i(t)^2 + \Delta_t^2}$. The spatio-temporal acceleration can be derived as the derivative of the velocity: $\hat{\mathbf{A}}_i(t) = (u_i(t) - u_i(t-1), v_i(t) - v_i(t-1), 0)$.
The curvature, following [15,17], is computed as

$$\hat{C}_i(t) = \frac{||\hat{\mathbf{V}}_i(t) \times \hat{\mathbf{A}}_i(t)||}{||\hat{\mathbf{V}}_i(t)||^3}. \tag{4}$$

The remaining two quantities are derived as $\hat{R}_i(t) = \frac{1}{\hat{C}_i(t)}$ and $\hat{A}_i(t) = \frac{\hat{V}_i(t)}{\hat{R}_i(t)}$.
Fig. 1 shows an example of the computed quantities for repetitive lifting actions. For the sake of clarity we focus on the trajectory of a single point (the centroid of the region, see Fig. 1(a)). In Fig. 1(b) the trend of the tangential velocity shows the presence of the well-known bell shape, typical of biological motion [14]. Notice the uneven level of noise in the features estimation: the velocity magnitude, directly measured from the optical flow, is the smoothest, while the other quantities, derived with further approximations, show a lower regularity.

4 Representing Biological Motion

Now that we have defined the low-level features, we may set up a procedure to describe and then classify the observed motion as instance of a biological or non-biological event.

At each time instant, we consider a regular grid of points for each region of interest segmented according to Sec. 3. With each of them we associate a feature vector following Eq. 3, and then combine their contributions comparing two different simple strategies, detailed in the following, reminiscent of possible coarse approaches to average the visual motion information.

4.1 Centroid-Based Descriptor

We first consider a coarse description obtained collapsing the whole information within a region $\mathcal{R}(t)$ in a single vector, i.e. the centroid, henceforth denoted (with an abuse of notation with respect to the previous use) as $\bar{\mathcal{F}}(\mathcal{R}(t)) = \bar{\mathcal{F}}$. Similarly to the original feature vectors, the centroid is a vector of heterogeneous features, that when compared should be appropriately handled. A way to deal with it is to normalize the data to a common range. A better alternative is to adopt a convex combination of kernel-based similarity functions, often referred to as Multi-Cue Integration in the supervised learning literature [21], and successfully applied to the problem of dynamic events modeling [16]. Let $\mathcal{R}$ and $\mathcal{R}'$ be two regions represented with their centroids $\bar{\mathcal{F}} = (\bar{V}, \bar{C}, \bar{R}, \bar{A})$ and $\bar{\mathcal{F}}' = (\bar{V}', \bar{C}', \bar{R}', \bar{A}')$. The Multi-Cue kernel $K_{MC} : \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}$ can be computed as the weighted sum of kernel-based functions $K : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ on each feature:

$$\begin{aligned} K_{MC}(\bar{\mathcal{F}}, \bar{\mathcal{F}}') = w_v K(\bar{V}, \bar{V}') + w_c K(\bar{C}, \bar{C}') + \\ w_r K(\bar{R}, \bar{R}') + w_a K(\bar{A}, \bar{A}') \end{aligned} \tag{5}$$

where the w's sum up to 1.

4.2 Histogram-Based Descriptor

We also consider a representation based on computing a histogram for each single feature, collecting the contributions of all points from a region. To this purpose, we first normalize each feature set so that all values are in the $[0 \ldots 1]$ range, then populate the 4 histograms and finally concatenate them to collect the final region descriptor. Henceforth, we will refer to the global region histogram as $\mathcal{H}(\mathcal{R}) = [\mathcal{H}_V \mathcal{H}_C \mathcal{H}_R \mathcal{H}_A]$.

Similarly to Sec. 4.1 we can treat each feature histogram independently, fusing their similarities in a single value while associating with them different weights. More formally, given two histograms $\mathcal{H}(\mathcal{R}) = \mathcal{H}$ and $\mathcal{H}(\mathcal{R}') = \mathcal{H}'$, a Multi-Cue kernel $K^{\mathcal{H}}_{MC} : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$, with M the total number of bin of the composed histogram, can be defined as

$$\begin{aligned} K^{\mathcal{H}}_{MC}(\mathcal{H}, \mathcal{H}') = w_v K^{\mathcal{H}}(\mathcal{H}_V, \mathcal{H}'_V) + w_c K^{\mathcal{H}}(\mathcal{H}_C, \mathcal{H}'_C) \\ w_r K^{\mathcal{H}}(\mathcal{H}_R, \mathcal{H}'_R) + w_a K^{\mathcal{H}}(\mathcal{H}_A, \mathcal{H}'_A) \end{aligned} \tag{6}$$

where $K^{\mathcal{H}} : \mathbb{R}^{\frac{M}{4}} \times \mathbb{R}^{\frac{M}{4}} \rightarrow \mathbb{R}$ is an appropriate measure to compare histograms.

5 Experimental Analysis

In this section we report the experimental analysis we conducted on a dataset acquired in-house. We structured the experimental analysis in two parts. On the first, we aim at validating the Two-Thirds Power Law in our setting, while evaluating the relative importance of each low-level feature we consider. On the second part, we focus instead on the biological motion classification problem, comparing the performances of the two descriptors introduced in Sec. 4 in combination with different kernels adopted in combination with SVM classifiers.

(a) Pointing (b) Lifting (c) Toy car (d) Bouncing ball

Fig. 2. Samples from the acquisitions of a subject from a single viewpoint (Fig. 2(a) and 2(b)), and of non biological motion events (Fig. 2(c) and 2(d)).

5.1 Data Set

We acquired indoor videos of two subjects observed from two slightly different viewpoints, performing repetitions of given actions from a *repertoire* of dynamic movements typical of an interaction setting, the one we have in mind. More in details, we consider *Gesticulating* while talking, *Pointing* a finger towards a certain 3D location (see Fig. 2(a)); *Waving* the hand from left to right and vice-versa; *Lifting* and object from the table to place it on a box (Fig. 2(b)); *Throwing* an object away; *Transporting* an object from and to different positions on the table. The latter is instantiated in two versions, with left-right and random object repositioning. Each video consists of 20 repetitions of the same atomic action (e.g. move the object from left to right); for each subject we acquired two videos in each view for each action, ending up with more than $20K$ frames.

As for the non-biological counterpart, we consider videos of a toy car (Fig. 2(c)), bouncing and rolling balls (Fig. 2(d)), a pendulum and a lever, for a total of about $10K$ data.

We split the set of videos in training set – used for model estimation – and test set – only adopted for performance evaluation. Model selection is based on K-fold cross validation with a grid search over the ranges of the parameters.

5.2 Proof of Concepts

On the Validity of the Two-Thirds Power Law. To assess the validity of the Two-Thirds Power Law for video analysis, we represent, for the sake of simplicity, the motion as a trajectory $\{\bar{\mathcal{F}}_t\}_{t=1}^{T}$ of centroids described according to Sec. 4.1. To correctly apply the law, we analyse the temporal sequences of their

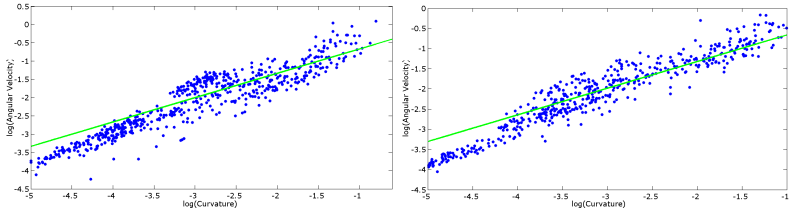


Fig. 3. Velocity *versus* curvature (log-log) measured on segments of trajectories describing sequences of lifting action performed by two different subjects.

velocity values, and then segment the trajectories in sub-parts considering portions between a maximum and a minimum in the sequence (dynamic instants in which the motion is subject to some variation, e.g. in acceleration or direction).

Following the seminal works [11,26], we analyse average velocity and curvature in each segments and show the obtained point in a log-log reference system. Fig. 3 reports two plots in which we collect observations from lifting actions performed by the two subjects in our dataset. They show a high correlation with the reference slope (i.e. $\frac{2}{3}$) in green.

We then fit each segment with an exponential function and estimate the exponents for both the biological and non-biological events. More in details, the average exponent for the biological population on the first view amounts to 0.65, and becomes 0.63 on the second view. A *two-sample t-test* confirms the high separation between average exponents for biological and non biological distributions (P-value < 0.0001).

On the Importance of the Features. In this section we investigate the relative importance of our motion features to characterize biological motion, despite their redundancy. To this purpose we consider a simple K-NN binary classifier and evaluate its accuracy for different feature vectors configurations – corresponding to using one or more features – and as the value of K increases. Since here we focus on the importance of each single feature of the vector, we adopt the centroid-based descriptor. To nullify the contribution of a feature we simply set to zero its weight in Eq. 5. From the results in Fig. 4(a) it is apparent the tangential velocity is the most relevant feature. The performances further increase when it is used in combination with other measures (see e.g. Fig. 4(b) and 4(c)).

A pros of the Multi-Cue Kernel is the fact that prior knowledge on the feature importance can be easily included in the model by appropriately tuning the weights. However, not always such information is available. An alternative is to learn the most appropriate weights from the data. We reported in Fig. 4(d) the weights selected as the best performing for increasing K values. There is a first range of Ks (from 1 to around 40) in which all the features are assigned an average importance, while for higher numbers of neighbors the curvature seems to be more relevant, but always if used in combination with some other information.

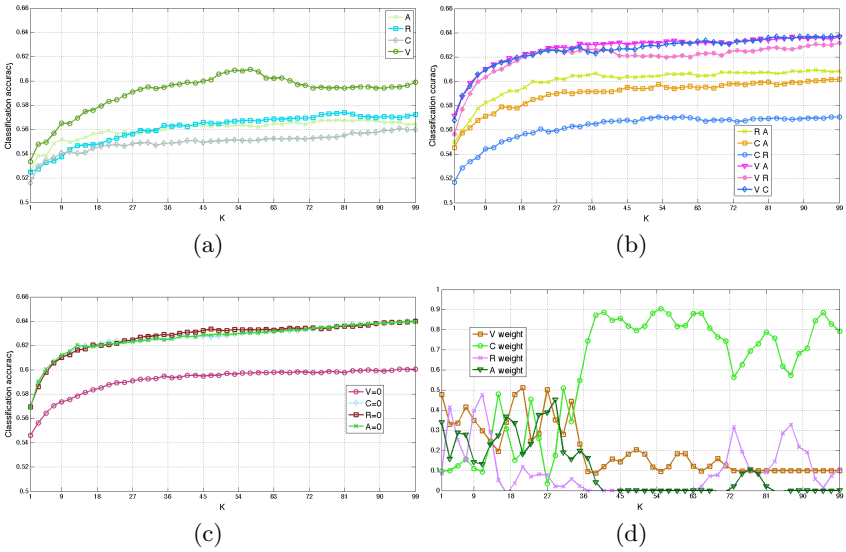


Fig. 4. An analysis of the relative importance of each feature.

To summarize, there is an empirical evidence of the relevance of all such features to the purpose of biological motion characterization. Since their *relative* importance may change depending on the specific category of human actions under analysis, the best option is to design an appropriate description by learning their importance (i.e. their weights) from the data. Nevertheless, the observation that all of them concur to best characterize our problem may be interpreted as a further evidence of the validity of the Two-Thirds Power Law: although relevant per-se, it is not the single feature that makes the difference, but its co-presence with the other measures, which are related to it by the law.

5.3 Experiments on Classification

We now focus more specifically on the problem of binary classification between biological and non-biological observations. To this end, we analyse the use of the two descriptors of Sec. 4 in combination with SVM classifiers.

Centroid-Based SVMs. We compare in the table of Fig. 5(a) the use of our instantaneous centroid-based description with different kernel functions, considering the mean accuracy computed on 5 different sampling of the input data set. As for the Multi-Cue similarities, we compare the case in which all the features are equally weighted with the values selected as best performing for some value of K using a K-NN on the training set (see previous section). The best performance is achieved with a Multi-Cue similarity function. We further test the ability of such kernel functions in classifying test data observed from the second

Kernel function	Acc.
Linear	56.66
Poly, $d = 2$	57.26
Poly, $d = 3$	56.90
Poly, $d = 4$	57.11
Radial basis, $\gamma = 0.1$	57.48
Sigmoid, $\gamma = 0.1$	55.01
(*) Multi-cue gaussian $\mathbf{w} = [0.25\ 0.25\ 0.25\ 0.25]$	65.42
(**) Multi-cue gaussian $\mathbf{w} = [0.5\ 0.2\ 0.2\ 0.1]$	66.17
(***) Multi-cue gaussian $\mathbf{w} = [0.1\ 0.7\ 0.1\ 0.1]$	64.41

(a)

Kernel function	Acc.
Linear	75.41
Histogram Inters.	75.69
Multi-Cue + Linear $\mathbf{w} = [0.5\ 0.2\ 0.2\ 0.1]$	76.37
Multi-Cue + Hist. Inters. $\mathbf{w} = [0.5\ 0.2\ 0.2\ 0.1]$	73.84
Multi-Cue + Gauss. $\mathbf{w} = [0.1\ 0.7\ 0.1\ 0.1]$	76.15

(b)

Fig. 5. Classification accuracy obtained with SVMs combined with different kernel methods. Left: using the centroid-based description. Right: using the histogram-based description.

viewpoint, obtaining 64.55 ± 1.54 for (*), 63.49 ± 2.46 for (**) and 64.3 ± 1.25 for case (***). Interestingly, the model is tolerant to viewpoint variation.

Furthermore, we may take into explicit account the temporal component by considering as input data series of temporally adjacent centroids. This requires and adaptation of the Multi-Cue function. Let $\mathcal{T} = [\bar{\mathcal{F}}_1 \dots \bar{\mathcal{F}}_T]$ and $\mathcal{T}' = [\bar{\mathcal{F}}'_1 \dots \bar{\mathcal{F}}'_T]$ be two sequences of centroids, then their Multi-Cue similarity is defined as

$$K_{MC}^{time}(\mathcal{T}, \mathcal{T}') = \sum_{t=1}^{T} K_{MC}(\bar{\mathcal{F}}_t, \bar{\mathcal{F}}'_t) \tag{7}$$

We consider as weights the best performing combination from the analysis of the single centroid (the one marked with (**)). We achieved the highest performace for $T = 20$, with accuracy 71.72 ± 1.45 on test data from view 1, and 65.49 ± 1.03 on test data from view 2 (training data are in both cases from view 1).

Histogram-Based SVMs. We conducted a similar analysis on the histogram-based descriptor, obtaining the performances reported on the table of Fig. 5(b). A first observation refers to the fact that the classification of instantaneous histograms outperforms the classification of centroids, even when they are supported by the temporal analysis. Also, the use of Multi-Cue kernel functions has a lower impact here, where the linear kernel is confirmed to be an appropriate choice, similarly to what happens in other classification problems built on top of histograms-like representations (see e.g. [2]). Even from a computational standpoint, the use of a linear kernel guarantees a high efficiency. The capability of handling a variation of the viewpoint is confirmed here, since the accuracy of classifying samples from view 2 using models trained on view 1 remains rather stable (76.02).

Extending the analysis to include temporal sequences of histograms (thus adapting the kernel, similarly to what done in Eq. 7) we obtain an accuracy of 89.03, which remains almost the same for view 2.

6 Final Discussion

In this paper we investigated the design of computational models of the primitives that are at the basis of the visual perception of motion in humans. Our inspiration roots on the very first stage of the human development, where the limited amount of visual information suggests that human beings have the capability of accomplishing certain perceptual tasks on the basis of rather coarse motion models. We took inspiration from the Two-Thirds Power Law, validating its applicability to video analysis problems. Moreover, we showed that a simple vector of low-level motion features, appropriately organized and handled in a learning framework, allows us to characterize biological motion against dynamic events due to non biological phenomena.

Our current investigations are devoted to the design of a hierarchical framework to replicate the developmental stages of human motion perception. On this respect, the capability of recognizing biological motion can be interpreted as the very first stage of such a system, to the purpose of localizing the possible target of interest before being able to interact with it.

A second stage in the refinement of human perception is the capability of understanding classes of actions, to focus on the important properties depending on the action. So, for manipulation actions, the relevant information may reside on the object. Alternatively, one may be interested on the environment, in presence of actions producing some kind of alteration on it. A preliminary investigation in this direction may be found in [15]. The aforementioned tasks set the scene for a more complete social awareness, that allows a subject to decode an action with respect to the final goal and the user intentions. For this task, more refined perception skills – and thus computational models – are required.

Acknowledgments. This research has been conducted in the framework of the European Project CODEFROR (FP7-PIRSES-2013-612555).

References

1. Casile, A., Giese, M.: Critical features for the recognition of biological motion. Jour. of Vision **5**, 348–360 (2005)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 2, pp. 886–893 (2005)
3. Falck-Ytter, T., Gredeback, G., von Hofsten, C.: Infants predict other people's action goals. Nature Neuroscience **9**(7), 878–879 (2006)
4. Fanello, S.R., Gori, I., Metta, G., Odone, F.: Keep it simple and sparse: Real-time action recognition. JMLR **14**(1), 2617–2640 (2013)
5. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 363–370. Springer, Heidelberg (2003)
6. Flach, R., Knoblich, G., Prinz, W.: The two-thirds power law in motion perception. Visual Cognition **11**(4), 461–481 (2004)

7. Hogan, N., Sternad, D.: On rhythmic and discrete movements: reflections, definitions and implications for motor control. Exp. Brain Res. **181**(1), 13–30 (2007)
8. Kanakogi, Y., Itakura, S.: Developmental correspondence between action prediction and motor ability in early infancy. Nat. Commun. **2**, 341 (2011)
9. Kandel, S., Orliaguet, J.P., Viviani, P.: Perceptual anticipation in handwriting: The role of implicit motor competence. Perc. and Psych. **62**(4), 706–716 (2000)
10. Kaplan, F., Hafner, V.: The challenges of joint attention. In: Int. Work. on Epigenetic Robotics (2006)
11. Lacquaniti, F., Terzuolo, C.: The law relating the kinematic and figural aspects of drawing movements. Acta Psychologica **54**, 115–130 (1983)
12. Lacquaniti, F., Terzuolo, C., Viviani, P.: The law relating the kinematic and figural aspects of drawing movements. Acta Psychologica **54**(13), 115–130 (1983)
13. Maoz, U., Portugaly, E., Flash, T., Weiss, Y.: Noise and the two-thirds power law. In: NIPS (2005)
14. Morasso, P.: Spatial control of arm movements. Experimental Brain Research **42**(2), 223–227 (1981)
15. Noceti, N., Sciutti, A., Rea, F., Odone, F., Sandini, G.: Estimating human actions affinities across views. In: VISAPP (2015)
16. Noceti, N., Odone, F.: Learning common behaviors from large sets of unlabeled temporal series. Image Vision Comput. **30**(11), 875–895 (2012)
17. Rao, C., Yilmaz, A., Shah, M.: View-invariant representation and recognition of actions. IJCV **50**(2), 203–226 (2002)
18. Richardson, M., Flash, T.: Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. Jour. of Neuroscience **22**(18), 8201–8211 (2002)
19. Sigala, R., Serre, T., Poggio, T.A., Giese, M.A.: Learning features of intermediate complexity for the recognition of biological motion. In: Duch, W., Kacprzyk, J., Oja, E., Zadrożny, S. (eds.) ICANN 2005. LNCS, vol. 3696, pp. 241–246. Springer, Heidelberg (2005)
20. Simion, F., Regolin, L., Bulf, H.: A predisposition for biological motion in the newborn baby. Proc. of the National Academy of Sciences **105**(2), 809–813 (2008)
21. Tommasi, T., Orabona, F., Caputo, B.: Discriminative cue integration for medical image annotation. PR Letters **29**(15) (2008)
22. Troje, N.F., Westhoff, C.: The inversion effect in biological motion perception: Evidence for a life detector? Current Biology **16**(8), 821–824 (2006)
23. Vieilledent, S., Kerlirzin, Y., Dalbera, S., Berthoz, A.: Relationship between velocity and curvature of a human locomotor trajectory. Neuroscience Letters **305**(1), 65–69 (2001)
24. Viviani, P., Baud-Bovy, G., Redolfi, M.: Perceiving and tracking kinesthetic stimuli: further evidence of motor-perceptual interactions. J. Exp. Psychol. Hum. Percept. Perform. **23**(4), 1232–1252 (1997)
25. Viviani, P., McCollum, G.: The relation between linear extent and velocity in drawing movements. Neuroscience **10**(1), 211–218 (1983)
26. Viviani, P., Stucchi, N.: Biological movements look uniform: evidence of motor-perceptual interactions. J. Exp. Psych. Hum. Perc. Perf. **18**(3), 603–623 (1992)
27. Viviani, P.: The relationship between curvature and velocity in two-dimensional smooth pursuit eye movements. Jour. of Neuroscience, 3932–3945 (1997)
28. Wang, X., Ma, X., Grimson, W.: Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. PAMI **31**(3), 539–555 (2009)